Industrial Crops and Products 58 (2014) 15–24

Industrial Crops and Products

journal homepage: www.elsevier.com/locate/indcrop

Comparison of response surface methodology (RSM) and artificial neural networks (ANN) towards efficient extraction of artemisinin from *Artemisia annua*

Josh L. Pilkington [a], Chris Preston [b], Rachel L. Gomes [a,*]

[a] *Faculty of Engineering, Manufacturing and Process Technologies Research Division, Department of Chemical and Environmental Engineering, University of Nottingham, University Park, Nottingham NG7 2RD, UK*

[b] *Bio Project Consulting Ltd., Ulverston, Cumbria LA12 9QL, UK*

ARTICLE INFO

Article history:
Received 11 October 2013
Received in revised form 2 March 2014
Accepted 18 March 2014
Available online 25 April 2014

Keywords:
Artemisia annua
Artemisinin
Artificial neural network
Response surface methodology
Extraction
Malaria

ABSTRACT

The solid-liquid extraction of *Artemisia annua* remains an important source of artemisinin, the precursor molecule to the most potent anti-malarial drugs available. Industrial manufacturers of artemisinin face many challenges in regards to volatile markets and sub-optimal extraction approaches. There is a need to improve current processing conditions, and one method is to model the processing options and identify the most appropriate process conditions to suit the market forces. This study examined the impact of extraction temperature, duration and solvent (petroleum ether) to leaf proportions on the recovery of artemisinin from leaf steeped in solvent, in a central composite design (CCD), and the results were used to generate both a response surface methodology (RSM) model and an artificial neural network (ANN) model.

Appraisal of the models through the coefficient of determination (R^2) and the absolute average deviation (AAD) showed that the ANN was superior (R^2 = 0.991, AAD = 1.37%) to the RSM model (R^2 = 0.903, AAD = 4.57%) in predicting artemisinin recovery. The ANN model was subsequently used to determine the optimal extraction conditions for the recovery of artemisinin, which were found to be an extraction duration of 8 h at a temperature of 45 °C and a leaf loading of 0.12 g/ml petroleum ether, from the conditions tested. An illustration is provided in how the results obtained from an ANN model may be used to determine optimal extraction conditions in response to market conditions. In addition, a co-solvency effect has been observed between extracted impurities and petroleum ether that substantially increases the solubility of artemisinin over that in petroleum ether alone, and which will require further investigation in the future. The impact of this co-solvency effect on the efficiency of artemisinin recovery in secondary extraction cycles was found to be significant.

1. Introduction

Artemisinin-based combination therapies (ACTs) are regarded by the World Health Organisation (WHO) as the most important class of drugs available in the fight against malaria, which claimed an estimated 660,000 lives in 2010 (WHO, 2013a). The critical starting material for all ACTs is artemisinin, which is primarily obtained through the solid-liquid extraction of the leaves of the *Artemisia annua* plant. An alternative source of artemisinin has recently been approved for use in ACTs (WHO, 2013b). This new source is from a genetically modified yeast that over produces artemisinic acid, which is then isolated and photocatalytically converted to artemisinin (Paddon et al., 2013). However, it is expected that the short to medium-term demand will still have to be met by *A. annua* extraction (A2S2, 2012). This leaves the artemisinin industry in a precarious position, with farmers disinclined to plant further crops, and industrial manufacturers facing numerous difficulties comprising volatile markets, variable biomass feedstock quality, unrecovered value in waste streams and technological limitations. There is a need for industrial manufacturers to optimise current approaches, thereby improving the profitability of production and ensuring a sufficient supply of artemisinin.

A manufacturing process cannot be optimised without first knowing the process details and such information is held by industry to retain market advantage. However, some heuristic rules can be provided for industry to review and apply for achieving their

* Corresponding author. Tel.: +44 115 846 8883.
E-mail address: rachel.gomes@nottingham.ac.uk (R.L. Gomes).

http://dx.doi.org/10.1016/j.indcrop.2014.03.016

own particular commercial objectives. Current industrial extractions use hexane, petroleum ether, toluene and HFC-134a, with petroleum ether being the most common solvent but demonstrating insignificant difference in performance to hexane (Christen and Veuthey, 2001; Lapkin et al., 2006; Vandenberghe et al., 1995). The solubility of artemisinin in a range of other solvents has been examined, which might lead to improved extraction of artemisinin (Lapkin et al., 2010; Liu et al., 2009; Nti-Gyabaah et al., 2010) due to their higher affinity to artemisinin, but they have not reached commercial application. This could be due to difficulties in the supply of sufficient quantities for extraction, increased cost implications and the increased risks associated with some solvents (acetonitrile, toluene, chloroform, etc.). The use of ethanol as an extraction solvent has been investigated by Fleming and Von Freyhold (2007) and, whilst positive results were obtained, the extracts are likely to contain higher quantities of sugars and polar impurities that hinder the subsequent crystallisation of artemisinin from the extract mixture (Lapkin et al., 2010). Other potential drawbacks to the industrial use of ethanol include a higher latent heat of vapourisation over petroleum ether that makes solvent recovery more expensive, in addition to the miscibility of ethanol and water that would result in the need for distillation to re-concentrate the ethanol after steam stripping of residual solvent in the extracted leaf bed.

Extraction is undertaken either by submerging the leaves in the extracting solvent (with or without agitation, which includes horizontal tumbling) or percolating the solvent through the leaf bed. The temperature range for extraction is 30–45 °C and the process may be undertaken over durations ranging from 8 to 48 h, with the possibility of additional extraction cycles to improve artemisinin recovery (Brisibe et al., 2008; Lapkin et al., 2006). Ethyl acetate may be added to the hexane/petroleum ether up to 5% by volume to increase the extent of artemisinin extraction (Brisibe et al., 2008) and to reduce the possibility of explosion through static build up and discharge (Lapkin et al., 2010). The proportion of leaf to solvent used by industry is difficult to ascertain from the literature, though it has been suggested that one kilogram of leaf could be extracted with one litre of solvent (Brisibe et al., 2008); such an approach would be impractical without the use of solvent percolation because the low density of dry biomass would ensure that the leaf bed greatly exceeds the level of extracting solvent.

With such a wide range of conditions reported in the literature, the task of process optimisation can only be undertaken using a methodology that can assess the individual impact of each process condition on overall efficiency. One such approach is the response surface methodology (RSM) developed by Box and Wilson (1951), which has seen wide application in the chemical industry due to its ability to optimise a process with a minimal amount of experimental data. As a statistical tool, RSM can model the impact of various process factors, both individually and through their cumulative interactions, on a system response, thereby providing an indication of the optimal operating region (Box et al., 2005). More recently, artificial neural networks (ANN) are finding increasing use as predictive tools in an extensive range of disciplines, including engineering, due to their ability to employ learning algorithms and discern input–output relationships for complex, nonlinear systems (Alavala, 2007; Zobel and Cook, 2011). Both RSM and ANN have been applied to optimise a range of natural product extraction processes and the resultant models show a strong correlation with experimental results. Recent examples include the extraction of phenols from mangosteen hull (Cheok et al., 2012), essential oils from *Diplotaenia cachrydifolia* (Khajeh et al., 2012), coumarin from *Cuscuta reflexa* (Mitra et al., 2011), secoisolariciresinol diglucoside from flaxseed (Nemes et al., 2012), oils from *Orthosiphon stamineus* (Pouralinazar et al., 2012) and *passiflora* seeds (Zahedi and Azarpour, 2011) and the extraction of natural dyes (Sinha et al., 2012, 2013).

The aim of this investigation was to develop and compare RSM and ANN models to indicate how the percentage recovery of artemisinin, through the extraction of *A. annua* using petroleum ether, might be optimised. The parameters investigated comprise solvent temperature, extraction duration and the proportion of leaf to solvent, in an extraction process that is considered to approximate the industrial approaches. The generated models were then compared in their suitability for predicting artemisinin recovery by analysing their coefficient of determination (R^2) and absolute average deviation (AAD) from experimental data. In the case of RSM, ANOVA was applied to assess any significant lack of fit with the experimental data. The models were then used to determine the impact of the extraction conditions on artemisinin recovery, thereby providing an indication of the optimal approach. The ANN model was then used to illustrate how processing conditions might be altered in order to optimise the first extraction cycle in response to market pressures. For the two case study conditions examined, an additional second extraction cycle was then performed to inform on the potential of improving the recovery of artemisinin further.

2. Materials and methods

2.1. Characterisation of A. annua

Samples of *A. annua* harvested from Wanzhou, Chongqing, China were supplied pre-milled (<2.80 mm) and dried by PIDI Standard (Holdings) Ltd (Guangzhou, Guangdong, China). The moisture content was determined to be 7.9 ± 0.2 wt% (n = 3) by drying to constant weight at 105 °C. The tapped bed density of the biomass was determined to be 0.21 ± 0.006 g/ml (n = 3) by filling and manually tapping a 50 ml measuring cylinder to provide an indication for the maximum solvent to leaf proportions that would be practical for extraction without solvent percolation.

The artemisinin content of the biomass was determined using a modified version of the method presented by Van Nieuwerburgh et al. (2006), using increased sample size and an extended extraction duration. Approximately 1.667 g of biomass was accurately weighted in triplicate and contacted with 10 ± 0.1 ml of chloroform (laboratory reagent grade, Fisher Scientific, UK) in 60 ml Boston type bottles, which were sealed and placed on a reciprocating shaker table operating at 170 rpm for a period of 5 min. After this duration, the supernatants were decanted and filtered to 0.2 μm using PTFE filter syringes (Fisher Scientific, Loughborough, UK), with 3 ml aliquots taken to dry down under atmospheric conditions (17 ± 1 °C). Filtration of artemisinin standard solutions using this methodology confirmed that there was no detectable decrease in artemisinin concentration due to adsorption onto the filter membrane. Prior to HPLC–UV analysis by a methodology used previously (Pilkington et al., 2012), the samples were reconstituted for a period of 24 h on a reciprocating shaker table operating at 170 rpm. The artemisinin content was found to be 1.37 ± 0.06 wt% of dry leaf.

2.2. Experimental design

The extraction parameters of solvent temperature (X_1), duration (X_2) and solvent to leaf proportions (X_3) were investigated for their impact on the recovery of artemisinin from *A. annua* using petroleum ether. Recovery is presented as the weight percentage of artemisinin detected in the extract mixture when compared to the total artemisinin present in the dry biomass. The temperature range of investigation was chosen to be 30–45 °C in accordance with the values published in the literature when hexane or petroleum

Table 1
The order and extraction conditions for the central composite design (CCD) including the coded levels of each parameter.

Run #	Extraction temp. (°C)	Duration (h)	Leaf concentration (g/ml)
1	37.5 (0)	6 (0)	0.118 (−α)
2	49.75 (+α)	6 (0)	0.2 (0)
3	25.3 (−α)	6 (0)	0.2 (0)
4	37.5 (0)	6 (0)	0.2 (0)
5	37.5 (0)	6 (0)	0.2 (0)
6	37.5 (0)	6 (0)	0.283 (+α)
7	37.5 (0)	9.266 (+α)	0.2 (0)
8	37.5 (0)	2.734 (−α)	0.2 (0)
9	30 (−1)	8 (+1)	0.15 (−1)
10	30 (−1)	8 (+1)	0.25 (+1)
11	37.5 (0)	6 (0)	0.2 (0)
12	45 (+1)	4 (−1)	0.15 (−1)
13	37.5 (0)	6 (0)	0.2 (0)
14	45 (+1)	8 (+1)	0.15 (−1)
15	37.5 (0)	6 (0)	0.2 (0)
16	30 (−1)	4 (−1)	0.15 (−1)
17	45 (+1)	8 (+1)	0.25 (+1)
18	30 (−1)	4 (−1)	0.25 (+1)
19	37.5 (0)	6 (0)	0.2 (0)
20	45 (+1)	4 (−1)	0.25 (+1)

ether are used as the extraction solvent (Brisibe et al., 2008; Lapkin et al., 2006). Published extraction durations of 8–48 h were considered too long to obtain an accurate representation of extraction dynamics due to the asymptotic gains from such extended extractions. It was therefore decided that extraction durations of 4–8 h would provide a more suitable basis for experiments, ensuring that the impact of duration could be more readily examined. For solvent to leaf proportions, it is clear that the dry biomass bed density of 0.21 g/ml (Section 2.1) would not allow for a ratio of 1 g leaf to 1 ml solvent without the use of solvent percolation. It is known to the authors that one industrial manufacturer used a ratio of 0.2 g/ml for their extractions and this value was therefore used as a starting basis. A range of 0.15–0.25 g/ml was selected for investigation, with 0.25 g/ml being possible due to the displacement of solvent by the leaf bed, which allows for all leaf to be submerged by the solvent level.

A central composite design (CCD) was selected to determine the experimental conditions as the inclusion of axial experimental points allow for a larger spread of conditions to be examined, which is beneficial when the required complexity of model is not known for accurate predictions to be made (Box et al., 2005). The three-factor experimental matrix was developed in Minitab® V.16, with 20 runs to include 8 factorial points, 6 centre points and 6 axial points. The resultant range of experimental conditions, including their coded levels ($-\alpha$,−1, 0, 1, α; $\alpha = 1.633$), can be observed in Table 1. The resultant recovery of artemisinin was then approximated by a quadratic equation (Eq. (1)):

$$R_{\text{art,pred}}(\%) = \beta_0 + \sum_{i=1}^{k} \beta_i X_i + \sum_{i=1}^{k} \beta_{ii} X_i^2 + \sum_{i<j} \sum \beta_{ij} X_i X_j + \varepsilon \qquad (1)$$

where $R_{\text{art,pred}}$ is the predicted overall recovery of artemisinin, β_0 is the constant coefficient, β_i, β_{ii} and β_{ij} are respectively the linear, quadratic and interaction coefficients, X_i and X_j are the independent variables and ε is the error. The statistical significance of each regression coefficient on the recovery of artemisinin was determined by analysis of variance (ANOVA).

2.3. Artificial neural network

Artificial neural networks (ANNs) are mathematical models that loosely approximate the function of biological neural networks. A multilayer perceptron (MLP) is a feed-forward ANN consisting of three or more layers of neurons, with the first layer of neurons representing the independent variable inputs. Each of the neurons in the first layer is connected to one or more layers of hidden neurons that represent nonlinear activation functions. These neurons are in turn connected to a final level of output neurons and, through the use of learning algorithms, the relative influence of each input neuron and their complex interactions on the observed result can be discerned.

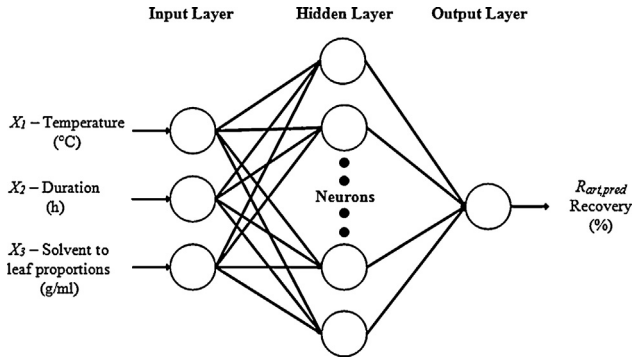

Fig. 1. Architecture of the developed artificial neural network (ANN).

An MLP was developed in MATLAB (The Mathworks, Inc., 2012a) with three input neurons representing the extraction temperature, duration and solvent to leaf proportions, a single hidden layer of neurons, and an output neuron representing the recovery of artemisinin. A representation of the MLP architecture can be observed in Fig. 1. The number of neurons required in the hidden layer was determined by trial and error to minimise the deviation of predictions from experimental results; a minimum of 10 neurons was required to build the final model utilising the data obtained to develop the RSM, and the addition of more neurons presented the possibility of over-fitting the model (Cheok et al., 2012; Madadlou et al., 2009). A total of 14 (70%) of experimental results were used to train the network, with the remaining results split evenly between network validation and testing. The ANN predictions were then used to generate surface and contour plots in SigmaPlot (Systat, V. 10.0).

2.4. Experimental procedure

Extractions were performed in 60 ml Boston type bottles (Fisherbrand, Fisher Scientific, UK), with temperature control (±0.3 °C) and agitation (170 rpm) provided by a linear shaking water bath (Fisherbrand, Fisher Scientific, UK). The required weight of leaf was measured into the 60 ml bottle, to which was added 25 ± 0.06 ml of petroleum ether (b.p. 60–80 °C; Analytical reagent grade, Fisher Scientific, UK) before being sealed and placed in the water bath for extraction to occur. After the specified duration, the extract was decanted directly without cooling and filtered to 0.2 μm by way of PTFE filter syringe (Fisher Scientific, Loughborough, UK). A 3 ml aliquot of each extract was then taken and reduced to dryness under atmospheric conditions (17 ± 1 °C) overnight in 7 ml vials. Once dry, 3 ml of acetonitrile was added to each vial, which were then sealed and suspended in an ultrasonic bath (VWR International, UK) rated at 160 W and frequency of 45 kHz for full disintegration of extract residues to occur, ensuring full reconstitution of artemisinin. The resulting solution was filtered to 0.2 μm through a PTFE filter syringe again to remove any insoluble particulates prior to HPLC–UV analysis by the methodology used previously (Pilkington et al., 2012).

Table 2
The experimentally obtained recovery of artemisinin compared to that predicted by the associated response surface methodology (RSM) model.

Run #	Experimental recovery (%)	RSM predicted recovery (%)
1	59.61	60.94
2	69.35	73.77
3	42.10	41.84
4	55.53	53.96
5	56.28	53.96
6	50.27	53.01
7	60.56	56.80
8	36.30	44.23
9	50.47	53.41
10	48.53	49.49
11	54.10	53.96
12	69.96	66.22
13	53.19	53.96
14	69.95	69.59
15	53.55	53.96
16	44.59	41.37
17	63.35	63.79
18	39.89	37.47
19	52.40	53.96
20	66.14	60.42

2.5. Process optimisation and secondary extraction cycles

From the model predictions, two sets of distinct extraction conditions (temperature, duration and leaf concentration) were used to illustrate how the first extraction cycle might be optimised to address external factors such as leaf availability, leaf cost and artemisinin market value. Fresh extractions of a different biomass batch from the same harvest location and season were then undertaken at both case study conditions following the procedure in Section 2.4.

Under laboratory conditions, the entrainment of extract in the leaf bed was found to be approximately 2.5 ml/g, which is greater than that expected in an industrial context due to their use of leaf bed presses for extract recovery. For this reason, immediately after recovering all available free extract from the first extraction, each leaf bed was contacted with 25 ± 0.06 ml of fresh petroleum ether (b.p. 60–80 °C; Analytical reagent grade, Fisher Scientific, UK) for 10 seconds at room temperature (17 ± 1 °C), with the wash solvent then being recovered in the same manner as the first extract and analysed by HPLC–UV for its artemisinin content. This served to dilute the remaining entrained mixture in the leaf bed prior to complete drying of the leaf bed at 45 ± 1 °C. The volume of recovered extract and subsequent petroleum ether wash was recorded in each instance.

The biomass was then subjected to a second extraction cycle under the same conditions as its respective first extraction cycle. For both the first and second extraction cycles, and the petroleum ether wash, the dried solvent residues were weighed prior to reconstitution in acetonitrile, to determine the weight proportion of artemisinin in each mixture.

3. Results and discussion

3.1. RSM model

The recovery of artemisinin predicted by the RSM model was compared to the experimental data and the results can be observed in Table 2, with the results plotted in Fig. 2 to provide the coefficient of determination (R^2 = 0.903). The coefficients for each term in the quadratic model are presented in Table 3 for un-coded units, and ANOVA was used to indicate which terms are statistically significant in the determination of recovery at a 95% confidence interval.

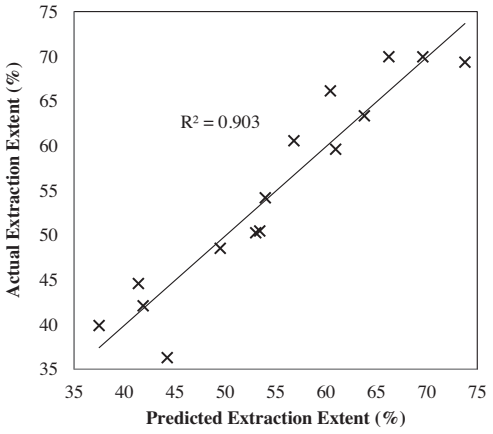

Fig. 2. Comparison between the artemisinin recovery predicted by the response surface methodology (RSM) model and the experimentally determined recovery (n = 20, with the 6 centre points combined as a mean value).

Table 3
Regression analysis of the response surface methodology (RSM) model for the recovery of artemisinin, with the associated statistical significance of each coefficient.

Coefficient	Value	F-value	P-value
β_0	3.3488	–	–
β_{X_1}	0.5084	67.24	0.000
β_{X_2}	11.2194	10.38	0.009
β_{X_3}	−179.6150	4.10	0.070
$\beta_{X_1X_1}$	0.0255	1.43	0.260
$\beta_{X_2X_2}$	−0.3231	1.16	0.306
$\beta_{X_3X_3}$	446.1780	0.89	0.367
$\beta_{X_1X_2}$	−0.1443	1.98	0.190
$\beta_{X_1X_3}$	−1.2600	0.09	0.765
$\beta_{X_2X_3}$	−0.0250	0.00	0.999

In general, it can be considered that higher Fisher's F-test values and lower P values indicate the relative significance of each term. It can be observed in Table 3 that the majority of the terms in the developed quadratic model are statistically insignificant (P > 0.05) when determining artemisinin recovery, and that the regression analysis is predominantly linear with respect to temperature and duration. Therefore the model was assessed for its suitability by examining the lack of fit through ANOVA, with the results presented in Table 4. It is clear from these results that the lack of fit for the quadratic model is significant due to the low probability (P = 0.004) of the Fisher's F-value for lack of fit, which is reinforced by the relatively low coefficient of determination (R^2 = 0.903) for the overall model.

An attempt was made to improve the accuracy of the model by removing insignificant terms from the quadratic equation. However, it was found that no combination of terms was able to improve the accuracy of the model, neither when excluding terms from the full quadratic equation nor when fresh RSM fitting was performed by neglecting the insignificant parameters. This result suggests that the variability in artemisinin recovery cannot be adequately predicted by the RSM model, taking into consideration the extraction temperature, duration and solvent to leaf proportions.

Table 4
Analysis of variance (ANOVA) to determine the suitability of the developed quadratic model in fitting the experimental data.

Source	Sum of squares	Deg. of freedom	F-value	P-value
Residual error	189.71	10	–	–
Lack-of-fit	178.93	5	16.59	0.004
Pure error	10.78	5	–	–

A possible explanation for this result is the complex relationship between artemisinin and co-extracts (such as oils and waxes), which can contribute to differences in artemisinin solubility (Lapkin et al., 2010). A large range of compounds have been identified in *A. annua* extracts (Bhakuni et al., 2001; Brown, 2010) and their extraction rates, both individually and in combination, are likely to influence the efficiency and rate of artemisinin extraction. This would be further compounded by the suggestion that extraction of metabolites could be somewhat sequential, with petroleum ether not acting as the extraction agent, but other groups of compounds that are initially extracted into petroleum ether that then serve to extract artemisinin (Lapkin et al., 2010). This suggestion is reinforced when examining the published data for the solubility of artemisinin in high purity hexane, which is stated to have comparable extraction properties to petroleum ether (Christen and Veuthey, 2001; Lapkin et al., 2006). The data provided by Nti-Gyabaah et al. (2010) has been used to plot the solubility across a range of temperatures with the results presented in Fig. 3.

It can be observed in Fig. 3 that artemisinin is only sparingly soluble in hexane across the range of temperatures examined in this study. However, by contrast, the range of artemisinin concentrations in the petroleum ether extracts generated in this study was 830–2050 μg/ml, with the artemisinin solubilised in solution even when extract samples were cooled to room temperature (17 ± 1 °C). This observation suggests that co-extracts could increase the solubility of artemisinin in petroleum ether based extracts by an order of magnitude.

Such interactions between artemisinin and co-extracts would be difficult to ascertain from RSM and would likely be batch-dependent anyway due to biomass variability, but the co-efficient of determination for the RSM model developed is sufficient to give overall indications on how the extraction process may be optimised. Fig. 4 illustrates the impact of the extraction variables on artemisinin recovery, with the third variable for each plot held at its high value (45 °C, 8 h or 0.25 g/ml). From Fig. 4(A) it can be observed that the influence of extraction duration on artemisinin recovery is likely to decrease as temperature is increased, indicating that not only can artemisinin be recovered more rapidly at higher temperatures, but also that significantly higher recoveries can be achieved than would otherwise be possible at lower temperatures by increasing the extraction duration. Fig. 4(B) indicates that the importance of extraction duration could diminish after approximately 7 h, suggesting that extended extraction durations could be decreased without significant loss of artemisinin recovery. The predicted decrease in artemisinin recovery with extraction

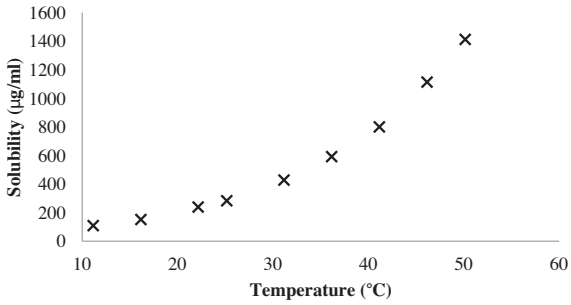

Fig. 3. The solubility of artemisinin in hexane at a range of temperatures using the data presented by Nti-Gyabaah et al. (2010).

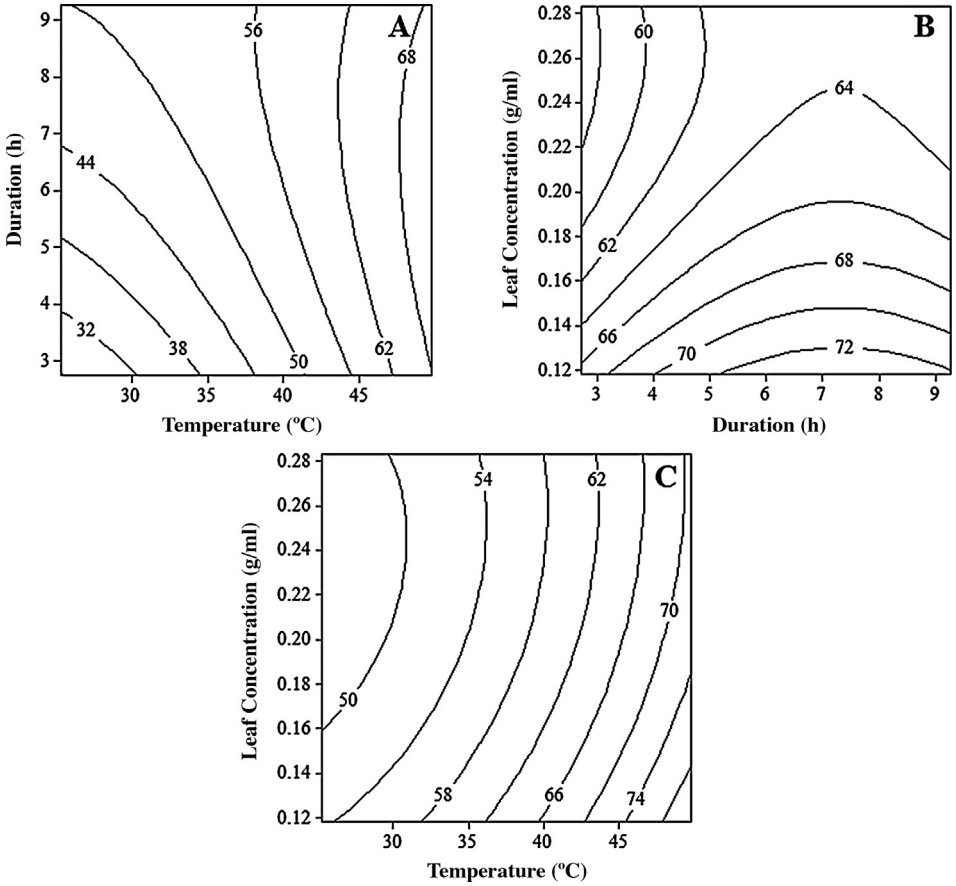

Fig. 4. Contour plots showing the influence of extraction parameters on the percentage recovery of artemisinin from *A. annua* as predicted by the RSM model with leaf concentration held constant at 0.25 g/ml (A), temperature held constant at 45 °C (B), and duration held constant at 8 h (C).

Table 5
The experimentally obtained recovery of artemisinin compared to that predicted by the associated artificial neural network (ANN) model.

Run #	Experimental recovery (%)	ANN predicted recovery (%)
1	59.61	60.46
2	69.35	69.35
3	42.10	42.10
4	55.53	54.18
5	56.28	54.18
6	50.27	50.27
7	60.56	60.65
8	36.30	36.30
9	50.47	50.47
10	48.53	48.53
11	54.10	54.18
12	69.96	69.96
13	53.19	54.18
14	69.95	69.95
15	53.55	54.18
16	44.59	44.59
17	63.35	62.18
18	39.89	37.36
19	52.40	54.18
20	66.14	62.92

durations beyond 7 h is counter-intuitive and is likely a result of the model parameters' lack of fit to experimental data. Finally, it is suggested in Fig. 4(C) that increasing the concentration of leaf in petroleum ether has a relatively minor influence on the recovery of artemisinin, indicating that higher biomass loading values may not have a detrimental impact on extraction efficiency, whilst allowing for improved productivity by increased biomass throughput.

3.2. ANN model

The recovery of artemisinin as predicted by the ANN is compared to the experimentally obtained values in Table 5. In order to test the suitability of the model, the predicted and actual results were plotted in Fig. 5 and the coefficient of determination ($R^2 = 0.991$) illustrates good agreement with the two sets of results.

By supplying the ANN model with matrices of extraction condition parameters, it was possible to visualise the relative impact of each extraction parameter using surface and contour plots generated in SigmaPlot (Systat, V. 10.0). Using the same approach as with the RSM model, the third variable for each plot was held at its high value (45 °C, 8 h or 0.25 g/ml) to generate the plots and the results of this investigation can be observed in Fig. 6. It is immediately obvious from Fig. 6(A) that, at extraction temperatures above approximately 40 °C, extraction duration has a negligible influence on the recovery efficiency of artemisinin as a state of equilibrium has been reached at the given temperature. In this circumstance (with leaf loading of 0.25 g/ml), the only way to increase the extraction of artemisinin is to increase the temperature of extraction, which is in line with the increasing solubility of artemisinin in extraction solvents with increasing temperature (Liu et al., 2009; Nti-Gyabaah et al., 2010). The results in Fig. 6(B) indicate that the effect of duration is generally neutral after 5 h, which suggests that the extraction has run to completion and that higher artemisinin recoveries could only be obtained by reducing the concentration of *A. annua* in the extraction solvent. This is in contradiction to the industrial approach suggested in the literature, with extended extraction times ranging from 8 to 48 h (Brisibe et al., 2008; Lapkin et al., 2006). Whilst reducing the leaf to solvent proportions can be seen to increase the recovery of artemisinin, an economic evaluation would need to be undertaken in order to assess the increased cost of solvent evaporation that would be required in order to induce crystallisation of artemisinin from the extract.

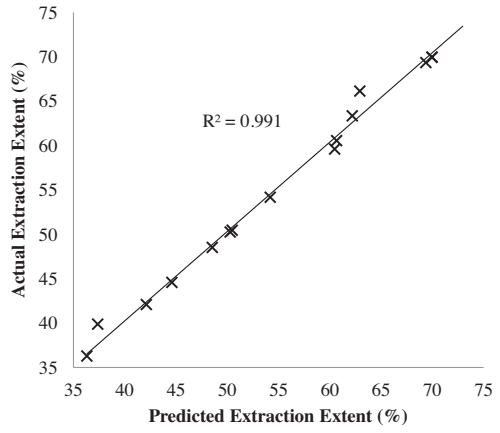

Fig. 5. Comparison between the artemisinin recovery predicted by the artificial neural network (ANN) model and the experimentally determined recovery ($n = 20$, with the 6 centre points combined as a mean value).

The results in Fig. 6(C) indicate an unusual trend at lower extraction temperatures, particularly between 25 and 30 °C. This is perhaps further evidence of a co-solvency effect between co-extracts and petroleum ether that are suggested in the literature (Lapkin et al., 2010). Increasing the leaf concentration from 0.12 g/ml to around 0.20 g/ml in this temperature range, and at an extraction duration of 8 h, leads to a decrease in the expected artemisinin recovery. However, a tipping point is reached, whereby the addition of further leaf serves to increase the artemisinin recovery. It is hypothesised that a critical concentration of one or more co-extracts is achieved and that the solubility of artemisinin is sufficiently increased in the extraction mixture as to promote increased recovery from the leaf. This is an important observation as it suggests that room temperature extractions with high leaf loading may be able to compete in terms of extraction efficiency with high temperature, low leaf loading extractions. As well as cost savings by running the extraction at lower temperatures, there is the potential that the extraction of certain impurities may be reduced, leading to a cleaner extract to be taken forward for purification. However, the maximum recovery of artemisinin in a single extraction cycle is only achieved with high temperatures and low leaf concentrations, and multiple extraction cycles or the addition of co-solvents would likely be required in order for low temperature, high loading extractions to be economically viable.

3.3. Comparison of RSM and ANN

Despite the lack of fit for the RSM model as determined in Section 3.1, it can still provide some indication of how the extraction process may be optimised when access to the necessary software to develop ANN architecture for a leaf batch is not available. For this reason, it is useful to determine the absolute average deviation (AAD) observed for both models to give an indication of how accurate the model predictions can be. The AAD is defined as in Eq. (2):

$$\mathrm{AAD}(\%) = \left(\frac{1}{n} \sum_{i=1}^{n} \left(\frac{R_{\mathrm{art,pred}} - R_{\mathrm{art,exp}}}{R_{\mathrm{art,exp}}} \right) \right) \times 100 \qquad (2)$$

where n is the number of sample points, $R_{\mathrm{art,pred}}$ is the predicted recovery of artemisinin and $R_{\mathrm{art,exp}}$ is the experimentally determined artemisinin recovery. The AAD for the RSM model was calculated to be 4.57%, whilst that of the ANN model was 1.37%. In addition to the coefficients of determination for both models ($R^2 = 0.991$ for ANN and $R^2 = 0.903$ for RSM), the AAD confirms that the ANN model is superior in predicting the recovery

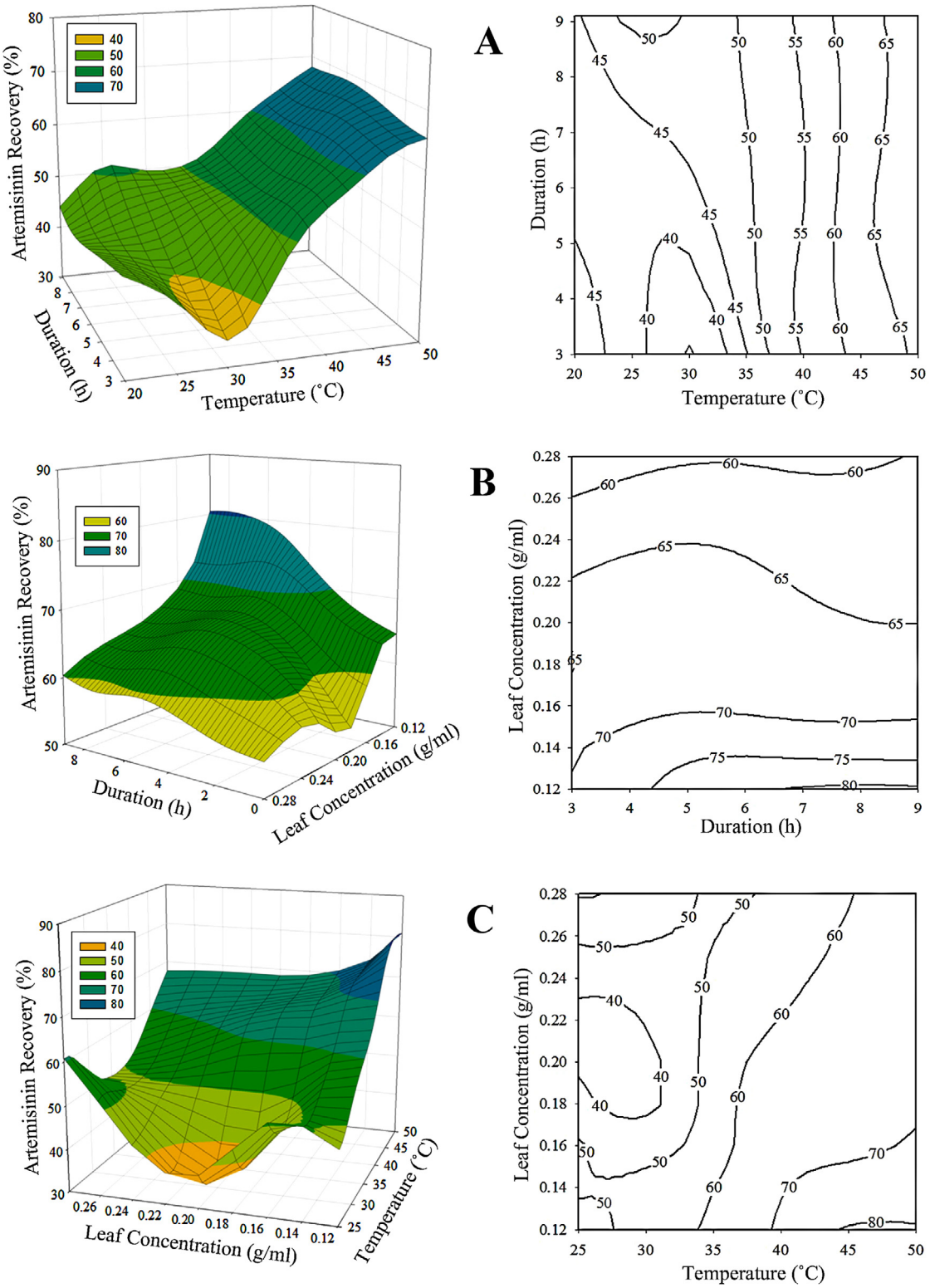

Fig. 6. Surface plots (left) and corresponding contour plots (right) showing the influence of extraction parameters on the percentage recovery of artemisinin from *A. annua* as predicted by the ANN model with leaf concentration held constant at 0.25 g/ml (A), temperature held constant at 45 °C (B), and duration held constant at 8 h (C).

of artemisinin from petroleum ether extraction, which is consistent with the findings of other research that compared both methodologies for natural product extraction (Cheok et al., 2012; Marchitan et al., 2010; Pouralinazar et al., 2012; Sinha et al., 2013; Zahedi and Azarpour, 2011). The ANN model has also identified an area in which co-solvency effects between petroleum ether and co-extracts may exist and have a significant impact on artemisinin recovery. This result is not unexpected but the interaction is not evident from the RSM model and this is the likely source of the resultant poor fit of the quadratic equation with experimental data. In order to address the poor fit, the factor levels would need to be revised to give better resolution around the local minima for

Table 6
Utilising the developed artificial neural network (ANN) to assist in determining the optimal artemisinin extraction conditions in a first-stage extraction when market conditions and/or *A. annua* supply is favourable (Case A) or restrictive (Case B).

Parameter	Case A	Case B
Temperature (°C)	45	45
Duration (h)	5	8
Leaf to solvent ratio (g/ml)	0.24	0.12
Leaf charged for extraction (kg)	960	480
Solvent use (L)	4000	4000
Leaf artemisinin content (wt%)	1	1
Artemisinin charged (kg)	9.6	4.8
Extraction efficiency (%)	65	80
Artemisinin recovered (kg)	6.24	3.84
Artemisinin loss to spent leaf (kg)	3.36	0.96
Output per day assuming 2 batches (kg)	12.48	7.68
Solvent cost	Standard	Increased
Leaf cost	Standard	Standard
Energy cost	Standard	Increased
Unit cost	Standard	Increased
Productivity	Standard	Reduced by 38%
Artemisinin wasted in leaf	Standard	Reduced from 35% to 20%

Table 7
The extraction of artemisinin from *Artemisia annua* with petroleum ether (b.p. 60–80 °C) under two experimental conditions in triplicate; Case A (45 °C, 5 h contact time and leaf concentration of 0.24 g/ml) and Case B (45 °C, 8 h contact time and leaf concentration of 0.12 g/ml). In each Case, the biomass was subjected to two sequential extraction cycles under the respective conditions with the artemisinin recovery and weight percentage of artemisinin reported for each stage, in addition to the combined extract properties.

Parameter	Case A	Case B
First extraction cycle		
Artemisinin recovery (%)	65.16 ± 0.89	75.94 ± 2.69
Proportion of artemisinin in extracted mixture (wt%)	17.43 ± 0.26	18.40 ± 0.39
Second extraction cycle		
Artemisinin recovery[a] (%)	34.73 ± 3.57	51.53 ± 4.10
Proportion of artemisinin in extracted mixture (wt%)	9.41 ± 0.86	10.32 ± 1.31
Combined extraction efficiency of both cycles		
Artemisinin recovery (%)	77.28 ± 0.87	87.84 ± 1.71
Proportion of artemisinin in extracted mixture (wt%)	13.59 ± 0.36	15.55 ± 0.09

[a] As a percentage of that remaining to be extracted from the biomass after the first extraction cycle.

artemisinin recovery, when the impact of temperature and leaf concentration at a fixed extraction duration is investigated. The ANN model does not suffer from this limitation and therefore is able to consider a wider range of processing conditions within a single experimental design.

3.4. Process optimisation and secondary extraction cycles

The results show that increased artemisinin recovery for a single extraction cycle is associated with elevated temperature, a low leaf to solvent ratio, and increased duration. In turn, these attributes all increase production costs by increased use of energy, reduced production plant throughput and increased solvent use. The optimum production process is that which is best aligned with these cost variables.

The developed ANN model in this study is only applicable to a first-stage extraction, and the extension of the model to multiple extraction cycles significantly increases the complexity of the system due to the increased number of variables. In particular, it is possible that maximum overall extraction efficiency would be achieved through a combination of different extraction conditions for the first, second and potentially third extraction cycles, necessitating prohibitively extensive experimental work to encompass all possible combinations. In the absence of such a model, a preliminary demonstration of how the existing model can be used as a decision making aid is presented in Table 6. The process described in Case A might be used when the availability of leaf is not limiting, and sales demand high productivity with good artemisinin sale price; Case B is a process that might be used when the supply of leaf is erratic or insufficient, and prices can compensate increased variable costs.

In light of the significant influence of *A. annua* co-extracts on the recovery of artemisinin, highlighted by the ANN model, it was prudent to then examine how the removal of impurities from leaf during the first extraction cycle might impact the performance of subsequent extraction cycles. For this purpose, the extraction conditions described for Case A and Case B in Table 6 were examined experimentally, with the addition of a second complete extraction cycle under the same conditions as their respective primary extraction cycles. The results from this investigation are presented in Table 7.

It can first be observed from Table 7 that for Case A, the recovery of artemisinin from the first extraction cycle was in agreement with the ANN model prediction (65%), whilst the experimental artemisinin recovery in Case B (75.94%) was lower than the predicted 80% under the extraction conditions examined. This was possibly due to the tested biomass necessarily coming from a similar but distinct batch (harvested at the same time but from a different location of the field) from that used to develop the ANN model, due to batch quantity limitations. In addition to the calculated recovery of artemisinin at these conditions, the results also indicate that approximately only 17–18% of the dry residue weight from the first extract mixture was artemisinin for both Cases.

In the instance of the second extraction cycle, it can be observed that both Case A and Case B demonstrated a significant decrease in the efficiency of artemisinin recovery. This finding supports the hypothesis that *A. annua* co-extracts are a significant driving force in the recovery of artemisinin, and that the removal of these components in the first extraction cycle impacts on the efficiency of subsequent extraction cycles. It should be noted that the mass balance for the second extraction cycle considers only the fresh artemisinin and impurities actively extracted from the leaf in that cycle, with any entrained quantities carried over from the previous leaf treatment being neglected. Such quantities could be recovered through washing of the leaf and not full extraction cycles.

An important observation is that the proportion of artemisinin to impurities actively extracted in the second extraction cycle under both experimental conditions is lower than the first cycle, indicating an increased burden on the subsequent purification stages. It was demonstrated by these results that there are still significant quantities of impurities remaining after the first extraction cycle, but that their presence has had a less profound impact on artemisinin recovery. A possible explanation for this result is that the type of impurities that are constructive in improving artemisinin recovery are mostly removed from the system after the first extraction cycle, leaving behind non-constructive impurities for the second extraction cycle. The interaction of the different impurities and their impact on the saturation levels of artemisinin in the extract mixture will therefore require far greater attention in future studies in order to make further progress in process optimisation.

The results in Table 7 highlight that process optimisation is not just a function of maximising the recovery of artemisinin, but involves a complex relationship between operating parameters and the application of a detailed cost-benefit analysis. It can be observed that the just one extraction cycle for Case B produced a higher-quality extract than two extraction cycles of Case A

combined; there is no significant difference in the eventual recovery of artemisinin in both cases, but the impurity burden in the primary extract of Case B is significantly reduced. An added advantage of Case B is that, despite a 50% lower leaf loading and longer extraction duration, the throughput of the processing plant would be increased due to the requirement of a single extraction cycle compared to the two extraction cycles of Case A. Careful consideration would be required before undertaking a second extraction for Case B as the recovery of approximately 12% more artemisinin from the biomass is accompanied by an increase in impurity concentrations, which may ultimately lead to a lower recovery post purification.

4. Conclusions

Despite recent advances in the semi-synthetic production of artemisinin, industrial solvent extraction from *A. annua* will remain an important source of the potent anti-malarial compound in the short to medium term. With industrial manufacturers facing increasing difficulties with market instability, there is a need to optimise the current extraction approaches. However, the range of extraction conditions stated in the literature is extensive, and the foundations for determining which approach to take in order to maximise processing efficiency are unavailable. Petroleum ether and hexane remain the most common solvents used for the extraction of artemisinin. Whilst other solvents have been identified in the literature that might improve the efficiency of extraction due to increased affinity for artemisinin over hexane/petroleum ether, considerations such as solvent cost, supply, higher boiling points (that increase solvent recovery costs) and increased safety concerns may have hampered their adoption.

In this study, the impact of extraction temperature, duration and solvent to leaf proportions was investigated on the extraction of artemisinin from *A. annua* using petroleum ether. In order to better understand the individual effects of the processing conditions, two modelling approaches were utilised, namely RSM and ANN which have extensive applications in process optimisation. This study found that an ANN model was superior to the RSM model both in terms of the coefficient of determination (R^2) and the absolute average deviation (AAD) when predicting the recovery of artemisinin. Utilising the ANN model to develop surface and contour plots of artemisinin recovery, it was found that significant co-solvency effects between co-extracts and petroleum ether exist at lower extraction temperatures. The resulting complex relationship between processing parameters and artemisinin recovery is likely to be the cause of the significant lack of fit observed in the RSM model. However, it is hypothesised that this co-solvency effect may be exploited to allow for low temperature, high leaf loading extractions to be undertaken without significant detrimental effects to processing efficiency. If such an approach is possible then cost savings could be made by industrial manufacturers, and the reduced temperature may serve to extract fewer impurities, leading to a reduced burden on subsequent purification stages. Extraction efficiencies exceeding 80% could not be achieved by the single steeping process used in this study.

Due to the significant impact that co-extracts have on the recovery of artemisinin, the effect of their removal in the first extraction cycle on the recovery of artemisinin in a subsequent batch extraction was examined. In the two case study examples investigated, it was found that the efficiency of the second extraction cycle was significantly reduced, supporting the hypothesis that the presence of impurities has a strongly positive influence on the recovery of artemisinin. In addition, the secondary extraction cycle was demonstrated to extract a higher proportion of impurities than the first cycle, indicating an increased burden on subsequent purification stages. Results from these studies indicate that process optimisation is a complex task of balancing extract quality with consideration of processing costs and plant throughput to compensate for changing market conditions.

Acknowledgements

The authors would like to acknowledge PIDI Standard (Holdings) Limited for donating the *A. annua* leaf required to undertake this study. The authors would also like to acknowledge the support of the Engineering and Physical Sciences Research Council (EPSRC) EP/J500483/1.

References

A2S2, 2012. Artemisinin Supply/Demand Forecast (updated November 2012). http://www.a2s2.org/market-data/a2s2-forecast.html (accessed 04.06.13).
Alavala, C.R., 2007. Fuzzy Logic and Neural Networks: Basic Concepts and Applications. New Age International, Daryaganj, Delhi.
Bhakuni, R.S., Jain, D.C., Sharma, R.P., Kumar, S., 2001. Secondary metabolites of *Artemisia annua* and their biological activity. Curr. Sci. 80, 35–48.
Box, G.E.P., Hunter, J.S., Hunter, W.G., 2005. Statistics for Experimenters: Design, Innovation, and Discovery, 2nd ed. Wiley-Interscience, Hoboken, NJ.
Box, G.E.P., Wilson, K.B., 1951. On the experimental attainment of optimum conditions. J. R. Stat. Soc. Ser. B Stat. Methodol. 13, 1–45.
Brisibe, E.A., Uyoh, E.A., Brisibe, F., Magalhaees, P.M., Ferreira, J.F.S., 2008. Building a golden triangle for the production and use of artemisinin derivatives against falciparum malaria in Africa. Afr. J. Biotechnol. 7, 4884–4896.
Brown, G.D., 2010. The biosynthesis of Artemisinin (Qinghaosu) and the phytochemistry of *Artemisia annua* L. (Qinghao). Molecules 15, 7603–7698.
Cheok, C.Y., Chin, N.L., Yusof, Y.A., Talib, R.A., Law, C.L., 2012. Optimization of total phenolic content extracted from *Garcinia mangostana* Linn. hull using response, surface methodology versus artificial neural network. Ind. Crop. Prod. 40, 247–253.
Christen, P., Veuthey, J.L., 2001. New trends in extraction, identification and quantification of artemisinin and its derivatives. Curr. Med. Chem. 8, 1827–1839.
Fleming, A., Von Freyhold, M., 2007. Assessing the technical and economic viability of the ethanolic extraction of Artemisia annua [online] http://www.mmv.org/newsroom/publications/assessing-technical-and-economic-viability-ethanolic-extraction-artemisia-annu.[accessed 04.06.13].
Khajeh, M., Moghaddam, M.G., Shakeni, M., 2012. Application of artificial neural network in predicting the extraction yield of essential oils of *Diplotaenia cachrydifolia* by supercritical fluid extraction. J. Supercrit. Fluids 69, 91–96.
Lapkin, A.A., Peters, M., Greiner, L., Chemat, S., Leonhard, K., Liauw, M.A., Leitner, W., 2010. Screening of new solvents for artemisinin extraction process using ab initio methodology. Green Chem. 12, 241–251.
Lapkin, A.A., Plucinski, P.K., Cutler, M., 2006. Comparative assessment of technologies for extraction of artemisinin. J. Nat. Prod. 69, 1653–1664.
Liu, Y., Iue, H., Pang, F., 2009. Solubility of artemisinin in seven different pure solvents from (283.15 to 323.15) K. J. Chem. Eng. Data 54, 762–764.
Madadlou, A., Emam-Djomeh, Z., Mousavi, M.E., Ehsani, M., Javanmard, M., Sheehan, D., 2009. Response surface optimization of an artificial neural network for predicting the size of re-assembled casein micelles. Comput. Electron. Agric. 68, 216–221.
Marchitan, N., Cojocaru, C., Mereuta, A., Duca, G., Cretescu, I., Gonta, M., 2010. Modeling and optimization of tartaric acid reactive extraction from aqueous solutions: a comparison between response surface methodology and artificial neural network. Sep. Purif. Technol. 75, 273–285.
Mitra, P., Barman, P.C., Chang, K.S., 2011. Coumarin extraction from *Cuscuta reflexa* using supercritical fluid carbon dioxide and development of an artificial neural network model to predict the coumarin yield. Food Bioprocess Technol. 4, 737–744.
Nemes, S.M., Orsat, V., Raghavan, G.S.V., 2012. Calibration of artificial neural network and partial least squares regression models for the prediction of secoisolariciresinol diglucoside contents in microwave-assisted extracts of various flaxseed (*Linum usitatissimum* L.) samples. Food Chem. 133, 1588–1595.
Nti-Gyabaah, J., Gbewonyo, K., Chiew, Y.C., 2010. Solubility of artemisinin in different single and binary solvent mixtures between (284.15 and 323.15) K and NRTL interaction parameters. J. Chem. Eng. Data 55, 3356–3363.
Paddon, C.J., Westfall, P.J., Pitera, D.J., Benjamin, K., Fisher, K., McPhee, D., Leavell, M.D., Tai, A., Main, A., Eng, D., Polichuk, D.R., Teoh, K.H., Reed, D.W., Treynor, T., Lenihan, J., Fleck, M., Bajad, S., Dang, G., Dengrove, D., Diola, D., Dorin, G., Ellens, K.W., Fickes, S., Galazzo, J., Gaucher, S.P., Geistlinger, T., Henry, R., Hepp, M., Horning, T., Iqbal, T., Jiang, H., Kizer, L., Lieu, B., Melis, D., Moss, N., Regentin, R., Secrest, S., Tsuruta, H., Vazquez, R., Westblade, L.F., Xu, L., Yu, M., Zhang, Y., Zhao, L., Lievense, J., Covello, P.S., Keasling, J.D., Reiling, K.K., Renninger, N.S., Newman, J.D., 2013. High-level semi-synthetic production of the potent antimalarial artemisinin. Nature 496, 528–532.
Pilkington, J.L., Preston, C., Gomes, R.L., 2012. The impact of impurities in various crude *A. annua* extracts on the analysis of artemisinin by liquid chromatographic methods. J. Pharmaceut. Biomed. 70, 136–142.

Pouralinazar, F., Yunus, M.A.C., Zahedi, G., 2012. Pressurized liquid extraction of *Orthosiphon stamineus* oil: experimental and modeling studies. J. Supercrit. Fluids 62, 88–95.

Sinha, K., Chowdhury, S., Das Saha, P., Datta, S., 2013. Modeling of microwave-assisted extraction of natural dye from seeds of *Bixa orellana* (Annatto) using response surface methodology (RSM) and artificial neural network (ANN). Ind. Crop. Prod. 41, 165–171.

Sinha, K., Das Saha, P., Datta, S., 2012. Response surface optimization and artificial neural network modeling of microwave assisted natural dye extraction from pomegranate rind. Ind. Crop. Prod. 37, 408–414.

Van Nieuwerburgh, F.C.W., Casteele, S.R.V., Maes, L., Goossens, A., Inze, D., Van Bocxlaer, J., Deforce, D.L.D., 2006. Quantitation of artemisinin and its biosynthetic precursors in *Artemisia annua* L. by high performance liquid chromatography–electrospray quadrupole time-of-flight tandem mass spectrometry. J. Chromatogr. A 1118, 180–187.

Vandenberghe, D.R., Vergauwe, A.N., Vanmontagu, M., Vandeneeckhout, E.G., 1995. Simultaneous determination of artemisinin and its bioprecursors in *Artemisia annua*. J. Nat. Prod. 58, 798–803.

WHO, 2013a. Acceptance of Non-Plant-Derived-Artemisinin Offers Potential to Increase Access To Malaria Treatment. http://apps.who.int/prequal/info_press/documents/pq_non-plant_derived_artemisinin_1.pdf (accessed 04.07.13).

WHO, 2013b. Malaria; Fact Sheet No. 94. http://www.who.int/mediacentre/factsheets/fs094/en/index.html (accessed 05/07/13).

Zahedi, G., Azarpour, A., 2011. Optimization of supercritical carbon dioxide extraction of Passiflora seed oil. J. Supercrit. Fluids 58, 40–48.

Zobel, C.W., Cook, D.F., 2011. Evaluation of neural network variable influence measures for process control. Eng. Appl. Artif. Intell. 24, 803–812.